

STN Columbus

* * * * * Welcome to STN International * * * * *

NEWS 1 Web Page URLs for STN Seminar Schedule - N. America
 NEWS 2 "Ask CAS" for self-help around the clock
 NEWS 3 May 12 EXTEND option available in structure searching
 NEWS 4 May 12 Polymer links for the POLYLINK command completed in REGISTRY
 NEWS 5 May 27 New UPM (Update Code Maximum) field for more efficient patent
 SDIs in Caplus
 NEWS 6 May 27 Caplus super roles and document types searchable in REGISTRY
 NEWS 7 Jun 28 Additional enzyme-catalyzed reactions added to CASREACT
 NEWS 8 Jun 28 ANTE, AQUALINE, BIOENG, CIVILENG, ENVIROENG, MECHENG,
 and WATER from CSA now available on STN(R)
 NEWS 9 Jul 12 BEILSTEIN enhanced with new display and select options,
 resulting in a closer connection to BABS
 NEWS 10 Jul 30 BEILSTEIN on STN workshop to be held August 24 in conjunction
 with the 228th ACS National Meeting
 NEWS 11 AUG 02 IFIPAT/IFIUDB/IFICDB reloaded with new search and display
 fields
 NEWS 12 AUG 02 Caplus and CA patent records enhanced with European and Japan
 Patent Office Classifications
 NEWS 13 AUG 02 STN User Update to be held August 22 in conjunction with the
 228th ACS National Meeting
 NEWS 14 AUG 02 The Analysis Edition of STN Express with Discover!
 (Version 7.01 for Windows) now available
 NEWS 15 AUG 04 Pricing for the Save Answers for SciFinder Wizard within
 STN Express with Discover! will change September 1, 2004
 NEWS EXPRESS JULY 30 CURRENT WINDOWS VERSION IS V7.01, CURRENT
 MACINTOSH VERSION IS V6.0c(ENG) AND V6.0Jc(JP),
 AND CURRENT DISCOVER FILE IS DATED 11 AUGUST 2004
 NEWS HOURS STN Operating Hours Plus Help Desk Availability
 NEWS INTER General Internet Information
 NEWS LOGIN Welcome Banner and News Items
 NEWS PHONE Direct Dial and Telecommunication Network Access to STN
 NEWS WWW CAS World Wide Web Site (general information)

Enter NEWS followed by the item number or name to see news on that
 specific topic.

All use of STN is subject to the provisions of the STN Customer
 agreement. Please note that this agreement limits use to scientific
 research. Use for software development or design or implementation
 of commercial gateways or other similar uses is prohibited and may
 result in loss of user privileges and other penalties.

* * * * * STN Columbus * * * * *

FILE 'HOME' ENTERED AT 18:02:58 ON 19 AUG 2004

=> file .pub

COST IN U.S. DOLLARS

SINCE FILE	TOTAL
ENTRY	SESSION
0.21	0.21

FULL ESTIMATED COST

FILE 'MEDLINE' ENTERED AT 18:03:13 ON 19 AUG 2004

STN Columbus

FILE 'BIOSIS' ENTERED AT 18:03:13 ON 19 AUG 2004
COPYRIGHT (C) 2004 BIOLOGICAL ABSTRACTS INC.(R)

=> s psiblast
L1 12 PSIBLAST

=> duplicate remove l1
DUPLICATE PREFERENCE IS 'MEDLINE, BIOSIS'
KEEP DUPLICATES FROM MORE THAN ONE FILE? Y/(N):n
PROCESSING COMPLETED FOR L1
L2 8 DUPLICATE REMOVE L1 (4 DUPLICATES REMOVED)

=> d 1-8 bib ab

L2 ANSWER 1 OF 8 MEDLINE on STN

Full Text

AN 2004328041 IN-PROCESS
DN PubMed ID: 15229883
TI Development and large scale benchmark testing of the PROSPECTOR_3
threading algorithm.
AU Skolnick Jeffrey; Kihara Daisuke; Zhang Yang
CS Center of Excellence in Bioinformatics, University at Buffalo, 901
Washington St., Suite 300, Buffalo, NY 14203, USA.. skolnick@buffalo.edu
NC GM-48835 (NIGMS)
SO Proteins, (2004 Aug 15) 56 (3) 502-18.
Journal code: 8700181. ISSN: 1097-0134.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS IN-PROCESS; NONINDEXED; Priority Journals
ED Entered STN: 20040702
Last Updated on STN: 20040722
AB This article describes the PROSPECTOR_3 threading algorithm, which
combines various scoring functions designed to match structurally related
target/template pairs. Each variant described was found to have a Z-score
above which most identified templates have good structural (threading)
alignments, Z(struct) (Z(good)). 'Easy' targets with accurate threading
alignments are identified as single templates with Z > Z(good) or two
templates, each with Z > Z(struct), having a good consensus structure in
mutually aligned regions. 'Medium' targets have a pair of templates
lacking a consensus structure, or a single template for which Z(struct) <
Z < Z(good). PROSPECTOR_3 was applied to a comprehensive Protein Data
Bank (PDB) benchmark composed of 1491 single domain proteins, 41-200
residues long and no more than 30% identical to any threading template.
Of the proteins, 878 were found to be easy targets, with 761 having a root
mean square deviation (RMSD) from native of less than 6.5 A. The average
contact prediction accuracy was 46%, and on average 17.6 residue
continuous fragments were predicted with RMSD values of 2.0 A. There were
606 medium targets identified, 87% (31%) of which had good structural
(threading) alignments. On average, 9.1 residue, continuous fragments
with RMSD of 2.5 A were predicted. Combining easy and medium sets, 63%
(91%) of the targets had good threading (structural) alignments compared
to native; the average target/template sequence identity was 22%. Only
nine targets lacked matched templates. Moreover, PROSPECTOR_3
consistently outperforms **PSIBLAST**. Similar results were predicted for
open reading frames (ORFs) < or =200 residues in the M. genitalium, E.
coli and S. cerevisiae genomes. Thus, progress has been made in
identification of weakly homologous/analogous proteins, with very high
alignment coverage, both in a comprehensive PDB benchmark as well as in
genomes.

L2 ANSWER 2 OF 8 MEDLINE on STN DUPLICATE 1
Full Text
 AN 2004107062 MEDLINE
 DN PubMed ID: 14997542
 TI Prediction of alpha-turns in proteins using PSI-BLAST profiles and secondary structure information.
 AU Kaur Harpreet; Raghava G P S
 CS Institute of Microbial Technology, Chandigarh, India.
 SO Proteins, (2004 Apr 1) 55 (1) 83-90.
 Journal code: 8700181. ISSN: 1097-0134.
 CY United States
 DT (EVALUATION STUDIES)
 Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 200404
 ED Entered STN: 20040304
 Last Updated on STN: 20040416
 Entered Medline: 20040415
 AB In this paper a systematic attempt has been made to develop a better method for predicting alpha-turns in proteins. Most of the commonly used approaches in the field of protein structure prediction have been tried in this study, which includes statistical approach "Sequence Coupled Model" and machine learning approaches; i) artificial neural network (ANN); ii) Weka (Waikato Environment for Knowledge Analysis) Classifiers and iii) Parallel Exemplar Based Learning (PEBLS). We have also used multiple sequence alignment obtained from **PSIBLAST** and secondary structure information predicted by PSIPRED. The training and testing of all methods has been performed on a data set of 193 non-homologous protein X-ray structures using five-fold cross-validation. It has been observed that ANN with multiple sequence alignment and predicted secondary structure information outperforms other methods. Based on our observations we have developed an ANN-based method for predicting alpha-turns in proteins. The main components of the method are two feed-forward back-propagation networks with a single hidden layer. The first sequence-structure network is trained with the multiple sequence alignment in the form of PSI-BLAST-generated position specific scoring matrices. The initial predictions obtained from the first network and PSIPRED predicted secondary structure are used as input to the second structure-structure network to refine the predictions obtained from the first net. The final network yields an overall prediction accuracy of 78.0% and MCC of 0.16. A web server AlphaPred (<http://www.imtech.res.in/raghava/alphapred/>) has been developed based on this approach.
 Copyright 2004 Wiley-Liss, Inc.

L2 ANSWER 3 OF 8 MEDLINE on STN
Full Text
 AN 2004233864 MEDLINE
 DN PubMed ID: 14594458
 TI PCAS--a precomputed proteome annotation database resource.
 AU Zhang Yong; Yin Yanbin; Chen Yunjia; Gao Ge; Yu Peng; Luo Jingchu; Jiang Ying
 CS College of Life Sciences, National Laboratory of Genetic Engineering and Protein Engineering, Center of Bioinformatics, Peking University, Beijing 100871, China.. zhangy@mail.cbi.pku.edu.cn
 SO BMC genomics [electronic resource], (2003 Nov 1) 4 (1) 42.
 Journal code: 100965258. ISSN: 1471-2164.
 CY England: United Kingdom

STN Columbus

DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 200406
 ED Entered STN: 20040511

Last Updated on STN: 20040615

Entered Medline: 20040614

AB BACKGROUND: Many model proteomes or "complete" sets of proteins of given organisms are now publicly available. Much effort has been invested in computational annotation of those "draft" proteomes. Motif or domain based algorithms play a pivotal role in functional classification of proteins. Employing most available computational algorithms, mainly motif or domain recognition algorithms, we set up to develop an online proteome annotation system with integrated proteome annotation data to complement existing resources. RESULTS: We report here the development of PCAS (ProteinCentric Annotation System) as an online resource of pre-computed proteome annotation data. We applied most available motif or domain databases and their analysis methods, including hmmpfam search of HMMs in Pfam, SMART and TIGRFAM, RPS-**PSIBLAST** search of PSSMs in CDD, pfscan of PROSITE patterns and profiles, as well as PSI-BLAST search of SUPERFAMILY PSSMs. In addition, signal peptide and TM are predicted using SignalP and TMHMM respectively. We mapped SUPERFAMILY and COGs to InterPro, so the motif or domain databases are integrated through InterPro. PCAS displays table summaries of pre-computed data and a graphical presentation of motifs or domains relative to the protein. As of now, PCAS contains human IPI, mouse IPI, and rat IPI, A. thaliana, C. elegans, D. melanogaster, S. cerevisiae, and S. pombe proteome. PCAS is available at <http://pak.cbi.pku.edu.cn/proteome/gca.php> CONCLUSION: PCAS gives better annotation coverage for model proteomes by employing a wider collection of available algorithms. Besides presenting the most confident annotation data, PCAS also allows customized query so users can inspect statistically less significant boundary information as well. Therefore, besides providing general annotation information, PCAS could be used as a discovery platform. We plan to update PCAS twice a year. We will upgrade PCAS when new proteome annotation algorithms identified.

L2 ANSWER 4 OF 8 MEDLINE on STN DUPLICATE 2
Full Text

AN 2002179124 MEDLINE

DN PubMed ID: 11911793

TI The efficient computation of position-specific match scores with the fast fourier transform.

AU Rajasekaran S; Jin X; Spouge J L

CS Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA.

SO Journal of computational biology : a journal of computational molecular cell biology, (2002) 9 (1) 23-33.

Journal code: 9433358. ISSN: 1066-5277.

CY United States

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 200206

ED Entered STN: 20020326

Last Updated on STN: 20020625

Entered Medline: 20020624

AB Historically, in computational biology the fast Fourier transform (FFT) has been used almost exclusively to count the number of exact letter matches between two biosequences. This paper presents an FFT algorithm that can compute the match score of a sequence against a position-specific

STN Columbus

scoring matrix (PSSM). Our algorithm finds the PSSM score simultaneously over all offsets of the PSSM with the sequence, although like all previous FFT algorithms, it still disallows gaps. Although our algorithm is presented in the context of global matching, it can be adapted to local matching without gaps. As a benchmark, our PSSM-modified FFT algorithm computed pairwise match scores. In timing experiments, our most efficient FFT implementation for pairwise scoring appeared to be 10 to 26 times faster than a traditional FFT implementation, with only a factor of 2 in the acceleration attributable to a previously known compression scheme. Many important algorithms for detecting biosequence similarities, e.g., gapped BLAST or **PSIBLAST**, have a heuristic screening phase that disallows gaps. This paper demonstrates that FFT algorithms merit reconsideration in these screening applications.

L2 ANSWER 5 OF 8 MEDLINE on STN

Full Text

AN 2001027874 MEDLINE
 DN PubMed ID: 10972829
 TI The spvB gene-product of the Salmonella enterica virulence plasmid is a mono(ADP-ribosyl)transferase.
 AU Otto H; Tezcan-Merdol D; Girisch R; Haag F; Rhen M; Koch-Nolte F
 CS Institute for Immunology, University Hospital, Martinistr. 52, D-20246 Hamburg, Germany.
 SO Molecular microbiology, (2000 Sep) 37 (5) 1106-15.
 Journal code: 8712028. ISSN: 0950-382X.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 200011
 ED Entered STN: 20010322
 Last Updated on STN: 20020420
 Entered Medline: 20001115
 AB A number of well-known bacterial toxins ADP-ribosylate and thereby inactivate target proteins in their animal hosts. Recently, several vertebrate ecto-enzymes (ART1-ART7) with activities similar to bacterial toxins have also been cloned. We show here that **PSIBLAST**, a position-specific-iterative database search program, faithfully connects all known vertebrate ecto-mono(ADP-ribosyl)transferases (mADPRTs) with most of the known bacterial mADPRTs. Intriguingly, no matches were found in the available public genome sequences of archaeabacteria, the yeast *Saccharomyces cerevisiae* or the nematode *Caenorhabditis elegans*. Significant new matches detected by **PSIBLAST** from the public sequence data bases included only one open reading frame (ORF) of previously unknown function: the spvB gene contained in the virulence plasmids of *Salmonella enterica*. Structure predictions of SpvB indicated that it is composed of a C-terminal ADP-ribosyltransferase domain fused via a poly proline stretch to a N-domain resembling the N-domain of the secretory toxin TcaC from nematode-infecting enterobacteria. We produced the predicted catalytic domain of SpvB as a recombinant fusion protein and demonstrate that it, indeed, acts as an ADP-ribosyltransferase. Our findings underscore the power of the **PSIBLAST** program for the discovery of new family members in genome databases. Moreover, they open a new avenue of investigation regarding salmonella pathogenesis.

L2 ANSWER 6 OF 8 MEDLINE on STN

Full Text

AN 2000497220 MEDLINE
 DN PubMed ID: 10972814
 TI DNase I homologous residues in CdtB are critical for cytolethal distending

STN Columbus

toxin-mediated cell cycle arrest.

AU Elwell C A; Dreyfus L A
 CS Division of Cell Biology and Biophysics, School of Biological Sciences,
 UMKC, Kansas City, MO 64110, USA.
 SO Molecular microbiology, (2000 Aug) 37 (4) 952-63.
 Journal code: 8712028. ISSN: 0950-382X.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 200010
 ED Entered STN: 20001027
 Last Updated on STN: 20001027
 Entered Medline: 20001019
 AB Cytotoxic distending toxins (CDTs) block cell division by arresting the
 eukaryotic cell cycle at G2/M. Although previously not recognized in
 standard BLAST searches, a position-specific iterated (PSI) BLAST search
 of the protein data bank using CDT polypeptides as query sequences
 indicated that CdtB bears significant position-specific homology to type I
 mammalian DNases. The **PSIBLAST** sequence alignment reveals that residues
 of DNase I involved in phosphodiester bond hydrolysis (His134 and His252)
 are conserved in CdtB as well as their respective hydrogen bond pairs
 (Glu78 and Asp212). CdtB also contains a pentapeptide motif found in all
 DNase I enzymes. Further, crude CDT preparations possess detectable DNase
 activity not associated with identical preparations from control cells.
 Five CdtB mutations in amino acids corresponding to DNase I active site
 residues were prepared and expressed together with wild-type CdtA and CdtC
 polypeptides. Mutation in four of the five DNase-specific active site
 residues resulted in CDT preparations that lacked DNase activity and
 failed to induce cellular distension or arrest division of HeLa cells.
 The fifth mutation, Glu86 (Glu78 in DNase I), retained the ability to
 induce a moderate level of cell cycle arrest and displayed reduced DNase
 activity relative to wild-type CDT. Together, these data suggest that the
 CDT holotoxin has intrinsic DNase activity that is associated with the
 CdtB polypeptide and that this DNase activity may be responsible for the
 CDT-induced cell cycle arrest.

L2 ANSWER 7 OF 8 MEDLINE on STN
 Full Text

DUPLICATE 3

AN 2001091283 MEDLINE
 DN PubMed ID: 11108697
 TI Ballast: blast post-processing based on locally conserved segments.
 AU Plewniak F; Thompson J D; Poch O
 CS Institut de Genetique et de Biologie Moleculaire et Cellulaire,
 Laboratoire de Biologie Structurale, (CNRS/INSERM/ULP), BP 163, 67404
 Illkirch Cedex, France.. plewniak@igbmc.u-strasbg.fr
 SO Bioinformatics (Oxford, England), (2000 Sep) 16 (9) 750-9.
 Journal code: 9808944. ISSN: 1367-4803.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 200101
 ED Entered STN: 20010322
 Last Updated on STN: 20010322
 Entered Medline: 20010125
 AB MOTIVATION: Blast programs are very efficient in finding relatively strong
 similarities but some very distantly related sequences are given a very
 high Expect value and are ranked very low in Blast results. We have
 developed Ballast, a program to predict local maximum segments (LMSs-i.e.

STN Columbus

sequence segments conserved relatively to their flanking regions) from a single Blast database search and to highlight these divergent homologues. The TblastN database searches can also be processed with the help of information from a joint BlastP search. RESULTS: We have applied the Ballast algorithm to BlastP searches performed with sequences belonging to well described dispersed families (aminoacyl-tRNA synthetases; helicases) against the SwissProt 38 database. We show that Ballast is able to build an appropriate conservation profile and that LMSs are predicted that are consistent with the signatures and motifs described in the literature. Furthermore, by comparing the Blast, **PsiBlast** and Ballast results obtained on a well defined database of structurally related sequences, we show that the LMSs provide a scoring scheme that can concentrate on top ranking distant homologues better than Blast. Using the graphical user interface available on the Web, specific LMSs may be selected to detect divergent homologues sharing the corresponding properties with the query sequence without requiring any additional database search.

L2 ANSWER 8 OF 8 MEDLINE on STN DUPLICATE 4
 Full Text
 AN 2000063280 MEDLINE
 DN PubMed ID: 10592246
 TI Assigning genomic sequences to CATH.
 AU Pearl F M; Lee D; Bray J E; Sillitoe I; Todd A E; Harrison A P; Thornton J M; Orengo C A
 CS Department of Biochemistry, University College London, University of London, Gower Street, London WC1E 6BT, UK.. frances@biochem.ucl.ac.uk
 SO Nucleic acids research, (2000 Jan 1) 28 (1) 277-82.
 Journal code: 0411011. ISSN: 0305-1048.
 CY ENGLAND: United Kingdom
 DT Journal; Article; (JOURNAL ARTICLE)
 LA English
 FS Priority Journals
 EM 200002
 ED Entered STN: 20000314
 Last Updated on STN: 20000314
 Entered Medline: 20000225
 AB We report the latest release (version 1.6) of the CATH protein domains database (<http://www.biochem.ucl.ac.uk/bsm/cath>). This is a hierarchical classification of 18 577 domains into evolutionary families and structural groupings. We have identified 1028 homo-logous superfamilies in which the proteins have both structural, and sequence or functional similarity. These can be further clustered into 672 fold groups and 35 distinct architectures. Recent developments of the database include the generation of 3D templates for recognising structural relatives in each fold group, which has led to significant improvements in the speed and accuracy of updating the database and also means that less manual validation is required. We also report the establishment of the CATH-PFDB (Protein Family Database), which associates 1D sequences with the 3D homologous superfamilies. Sequences showing identifiable homology to entries in CATH have been extracted from GenBank using PSI-BLAST. A CATH-**PSIBLAST** server has been established, which allows you to scan a new sequence against the database. The CATH Dictionary of Homologous Superfamilies (DHS), which contains validated multiple structural alignments annotated with consensus functional information for evolutionary protein superfamilies, has been updated to include annotations associated with sequence relatives identified in GenBank. The DHS is a powerful tool for considering the variation of functional properties within a given CATH superfamily and in deciding what functional properties may be reliably inherited by a newly identified relative.

STN Columbus

=> log y

COST IN U.S. DOLLARS

SINCE FILE

TOTAL

ENTRY

SESSION

FULL ESTIMATED COST

2.99

3.20

STN INTERNATIONAL LOGOFF AT 18:03:54 ON 19 AUG 2004